Amendments to the Specification:

Please replace the three paragraphs beginning at page 1, line 4 with the following paragraphs.

This application is a continuation of "METHODS OF POPULATING DATA STRUCTURES FOR USE IN EVOLUTIONARY SIMULATIONS" by Selifonov and Stemmer, USSN PCT/US00/01138, Filed January 18, 2000, and is a continuation-in-part of "METHODS OF POPULATING DATA STRUCTURES FOR USE IN EVOLUTIONARY SIMULATIONS" by Selifonov and Stemmer, USSN 09/416,837 (now abandoned), filed October 12 1999.

This application is also a continuation-in-part of "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES AND POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov et al., filed January 18, 2000, USSN 09/494,282, and is a continuation-in-part of "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES AND POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov et al., filed January 18, 2000, USSN PCT/US00/01202; which are continuation-in-part applications of "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES AND POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov et al., USSN 09/416,375 (now abandoned), filed October 12, 1999, which is a non provisional of "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES AND POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov and Stemmer, USSN 60/116,447, filed January 19, 1999 and a non-provisional of "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES AND POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov and Stemmer, USSN 60/118,854, filed February 5, 1999.

This application is also a continuation-in-part of "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" by Crameri et al., filed January 18, 2000, USSN 09/484,850 (now US Patent No. 6,368,861) and a continuation-in-part of "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" by Crameri et al., filed January 18, 2000, USSN PCT/US00/01203, which are continuation-in-part applications of "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID

RECOMBINATION" by Crameri et al., USSN 09/408,392 (now US Patent No. 6,376,246), filed September 28, 1999, which is a non-provisional of "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" by Crameri et al., USSN 60/118,813, filed February 5, 1999 and a non-provisional of "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" by Crameri et al., USSN 60/141,049, filed June 24, 1999.

P.O. Box 778
Berkeley, CA 94704-0778

Please replace the paragraph beginning at page 8, line 7 with the following paragraph:

The term "data structure" refers to the organization and optionally associated device for the storage of information, typically multiple "pieces" of information. The data structure can be a simple recordation of the information (*e.g.*, a list) or the data structure can contain additional information (*e.g.*, annotations) regarding the information contained therein, can establish relationships between the various "members" (information "pieces") of the data structure, and can provide pointers or linked to resources external to the data structure. The data structure can be intangible but is rendered tangible when ~~be~~ stored/represented in tangible medium. The data structure can represent various information architectures including, but not limited to simple lists, linked lists, indexed lists, data tables, indexes, hash indices, flat file databases, relational databases, local databases, distributed databases, thin client databases, and the like. In preferred embodiments, the data structure provides fields sufficient for the storage of one or more character strings. The data structure is preferably organized to permit alignment of the character strings and, optionally, to store information regarding the alignment and/or string similarities and/or string differences. In one embodiment this information is in the form of alignment "scores" (*e.g.*, similarity indices) and/or alignment maps showing individual subunit (*e.g.*, nucleotide in the case of nucleic acid) alignments. The term "encoded character string" refers **to** a representation of a biological molecule that preserves desired sequence/structural information regarding that molecule.

Please replace the paragraph beginning at page 14, line 27 with the following paragraph:

The biological molecule(s) are encoded into character strings. In the simplest instance, the character string is identical to the character code used to represent the biological molecule. Thus, for example, the character string can comprise the characters A, C, G. T, or U where a nucleic acid is encoded. Similarly, the standard amino acid nomenclature can be used to represent a polypeptide sequence. Alternatively, it will be realized that, to some extent, the encoding scheme **is** arbitrary. Thus, for example, in the case of nucleic acids the A, C, G, T, or U can be represented by the integers 1, 2, 3, 4, and 5, respectively and the nucleic acid can be represented as a string of these integers which is itself a single (albeit typically large) integer. Other coding schemes are also possible. For example, the biological molecule can be encoded into a character string where each "subunit" of the molecule is encoded into multi-character representation. Alternatively various compressed representations are also possible (*e.g.*, where recurrent motifs are represented only once with appropriate pointers identifying each occurrence).

Please replace the paragraph beginning at page 15, line 19 with the following paragraph:

In a preferred embodiment, the character string encoded biological molecules provide an initial population of strings from which substrings are selected. Typically at least two[[,]] substrings are selected with one substring coming from each initial character string. Where there are more than two initial character strings, it is not necessary that every initial character string provide a substring as long as at least two initial character strings provide such substrings. In preferred embodiments, however, at least one substring will be selected from each initial string.

P.O. Box 778
Berkeley, CA 94704-0778

Please replace the paragraph beginning at page 16, line 7 with the following paragraph:

Preferred substrings are also selected so as to not be unduly short. Typically a substring will be no shorter than the ~~minim~~ minimum string length necessary to represent one subunit of the encoded biological molecule. Thus, for example, where the encoded biological molecule is a nucleic acid the substring will be long enough to at least encode one nucleotide. Similarly, where the encoded biological molecule is a polypeptide the substring will be long enough to at least encode one amino acid.

P.O. Box 778
Berkeley, CA 94704-0778

Please replace the paragraph beginning at page 23, line 3 with the following paragraph:

In one embodiment, the substrings are randomly concatenated to produce "recombined" strings. In one approach to such "random" concatenation, each substring is assigned a unique identifier (*e.g.,* an integer or other identifier). The identifiers are then randomly selected from the pool (*e.g.,* using a random number generator) and the subsequences corresponding to those identifiers are joined to produce a concatenated sequence. When joined subsequences are approximately ~~ore~~ **or** exactly the length of the starting character string(s), the process is started anew to produce another string. The process is repeated until all of the substrings are utilized. Alternatively the substrings can be selected without withdrawing them from the "substring pool" and the process is repeated until a desired number of "full-length" strings are obtained.

P.O. Box 778
Berkeley, CA 94704-0778

Please replace the paragraph beginning at page 24, line 11 with the following paragraph:

It will not ~~required~~ **require** that perfect order be established in every concentrated character string. That a percentage (*e.g.*, preferably at least 1 percent, more preferably at least 10 percent, still more preferably at least 20% and most preferably at least 40 percent, at least 60% or at least 80%) of the concatenated sequences preserve the original order is preferred.

Please replace the paragraph beginning at page 25, line 17 with the following paragraph:

In one embodiment, a similarity index can be used as a selection criterion. Thus newly generated concatenated character strings must share a particular ~~predefinined~~ **predefined** similarity (*e.g.*, greater than 10%, preferably greater than 20% or 30%, more preferably greater than 40% or 50% and most preferably greater than 60%, 70%, 80%, or even 90%) with each other and/or with the initial strings (or the encoded molecules) and/or with ~~a~~ one or more "reference" strings.

Please replace the paragraph beginning at page 35, line 13 with the following paragraph:

Methods of implementing Intranet and/or Intranet embodiments of computational and/or data access processes are well known to those of skill in the art and are ~~documentede~~ **documented** in great detail (*see e.g.,* Cluer *et al.* (1992) *A General Framework for the Optimization of Object-Oriented Queries, Proc SIGMOD International Conference on Management of Data, San Diego, California,* Jun 2-5, 1992, SIGMOD Record, vol. 21, Issue 2, Jun., 1992; Stonebraker, M., Editor ACM Press pp. 383-392; ISO-ANSI, Working Draft, "Information Technology-Database Language SQL", Jim Melton, Editor, International Organization for Standardization and American National Standards Institute, Jul. 1992; Microsoft Corporation, "ODBC 2.0 Programmer's Reference and SDK Guide. The Microsoft Open Database Standard for Microsoft Widows.TM. and Windows NT.TM., Microsoft Open Database Connectivety.TM. Software Development Kit", 1992, 1993, 1994 Microsoft Press, pp. 3-30 and 41-56; ISO Working Draft, "Database Language SQL-Part 2: Foundation (SQL/Foundation)", CD9075-2:199.chi.SQL, Sep. 11, 1997, and the like).

Please replace the paragraph beginning at page 40, line 14 with the following paragraph:

In contrast, the collections of character strings ~~produces~~ **produced** by the methods of this invention contain far more information than the randomly produced starting points used in conventional evolutionary algorithms. First, each member of population contains considerable information regarding molecular structure. Thus, one member is distinguished from another member not simply as "self/not-self" (*i.e.,* al allelic representation), but rather members are distinguished by degrees of relatedness/similarity. Members of the populations produced by the methods of this invention will reflect varying degrees of covariation.

Please replace the paragraph beginning at page 40, line 31 with the following paragraph:

In another embodiment, the data structures generated by the methods of this invention can be used as tags (indices) for indexing essentially any kind of information. ~~IN~~ **In** this approach, information of greater similarity is tagged using members of the data structure (character strings) having greater similarity, while information of lower similarity is tagged with members of the data structure having lower similarity. In preferred embodiments, the similarity of the character strings used to tag two different pieces of data reflects (is proportional to) the similarity of tagged information.

Please replace the paragraph beginning at page 42, line 20 with the following paragraph:

If members of the data structure are tested under a specific set of conditions for a particular property, the optimal combinations of sequences from the data structure (or the initial string collection) for those conditions can be determined. If the assay conditions are altered in only one parameter, different individuals from the library (data structure) will be identified as the best performers. Because the screening conditions are very similar, most amino acids will probably be conserved between the two sets of best performers (the best performers in the initial string collection (set 1) and the best performers in the populated data structure (set 2)). Comparisons of the sequences of ~~t~~ **the** best enzymes under the two different conditions will therefore identify the sequence differences responsible ~~of~~ **for** the differences in performance.